Amendments to the Specification:

Please replace paragraph [077] with the following amended paragraph:

Significant signals are also associated with particular structural features of proteins. Therefore, the presence and type of significant signals within an amino acid sequence can be used to predict structural features of a protein having the amino acid sequence. Significant signals show conservation between related proteins, for example, cognate proteins from different species. The identification of conserved significant signals between proteins can therefore be used to identify conserved structural features of the proteins, and therefore which segments of the proteins are critical for function. The conserved segments of proteins identified by conservation of significant signals are not coextensive with conserved segments identified by primary sequence analysis. Therefore, the described methods detect conserved regions of proteins that are missed by conventional approaches. For example, if the predetermined set of amino acids consists of C, I, L, M, F, W, Y and V, the following three amino acid sequences all generate the same signal (001011101) despite the fact that they contain no sequence identity: ANISVYYEM (SEQ ID NO:1); TSFNFWMGV (SEQ ID NO:2); SGCGLILNC (SEQ ID NO:3). Three proteins that have these respective sequences at a particular position do not demonstrate amino acid similarity but their signal designations are identical. Significant signals that generate specific structural features of proteins can therefore predict common structural features between proteins without the need to rely on amino acid or nucleotide similarity to detect such features.

Please replace paragraph [095] with the following amended paragraph:

As an example, the amino acid sequence PAGEQEAFPPN (SEQ ID NO:4) has 3 window lengths of 9. Transformed into a binary code according to the class 1 amino acid set, the sequence reads 00000000100. Using the designations mentioned in the above paragraph, the sequence of designations for this amino acid sequence using a window length of 9 reads 002-003-004. Any protein sequence can be transformed into a sequence of designations for any set of amino acids.

Please replace paragraph [0151] with the following amended paragraph:

We used a binary signal model in which each of the 20 amino acids was assigned a value of 0 or 1. We defined signals as the pattern of 1's that appears in protein sequences when transformed using the model. For example, consider the ARQELKM (SEQ ID NO:5) amino acid set. A protein sequence was transformed by assigning a 1 to all residues that are members of the set of those seven amino acids, and assigning a 0 to all other residues. If we used a sequence window nine residues in length, then there are a total of 2^9, or 512 different possible signals. The signal strength for each signal, N_{ss}, is the number of selected amino acids in the particular signal, or equivalently the sum of the transformed digits. For example, the signal 011011100 has a signal strength of 5.

Please replace paragraph [0152] with the following amended paragraph:

If binary signals exist in protein sequences then we expected to find linguistic structure in the sequences. One way to detect such structure is to compare the actual signal strength distribution with the expected distribution if protein sequences were random. For a given sequence window length, N_w, we scanned our sequence database to determine the distribution of the $N_w + 1$ signal strength values. We then used the binomial distribution to compute the signal strength frequencies in random protein sequences. The binomial distribution is a function of N_w and the abundance of the selected amino acids, f_{aa}. For the ARQELKM (SEQ ID NO:5) amino acid set, f_{aa} is 0.397 in our collection of 790 protein sequences. Figure 1 shows the actual and random signal strength distributions for the ARQELKM (SEQ ID NO:5) amino acid set.

Please replace paragraph [0153] with the following amended paragraph:

Figure 1 shows that the amino acids ARQELKM (SEQ ID NO:5) tend to cluster with respect to random sequences. That is, in a sequence segment of nine amino acids, the ARQELKM (SEQ ID NO:5) amino acids, taken together as group of like monomers, tend to appear more often in either low or high numbers (0-2 and 6-9) and less often in medium numbers (3-5).

Please replace paragraph [0159] with the following amended paragraph:

Both of our optimization methods for searching for χ^2 local maxima led to the same results. We found two useful amino acids sets with a χ^2 local maximum, ARQELKM (SEQ ID NO:5) and CILMFWYV (SEQ ID NO:6). There also exist two other redundant, identical χ^2 local maxima corresponding to the respective complementary amino acids sets. Figure 2 shows the actual and random signal strength distributions for the useful CILMFWYV (SEQ ID NO:6) amino acid set.

Please replace paragraph [0160] with the following amended paragraph:

Figure 2 shows that the CILMFWYV (SEQ ID NO:6) amino acid set tends to anticluster with respect to random sequences. The set has lower frequencies in the extreme signal strength values (0-1 and 5-9) and higher frequencies in the middle signal strength values (2-4). In this case the χ^2 value is 5,173 and the probability that the parent distribution is random is 10^{-1114}.

Please replace paragraph [0161] with the following amended paragraph:

The signal 001100100 for the CILMFWYV (SEQ ID NO:6) amino acid set is a statistically significant signal as it occurs 801 times in our database but would be expected to occur only 479 times in random sequences of equal length, according to Equation 2. The signal frequency is therefore 801/479, or 1.67. The signal frequency may be sub or super unity, and statistically significant signals may have low or high frequencies. For this reason it is also useful to compute the corresponding sequence χ^2 value. This single category in the χ^2 calculation is a useful metric of the statistical significance of the signal's occurrences in actual protein sequences. For this signal the sequence χ^2 value is 216.3.

//

Please replace Table 1 with the following amended Table 1:

Table 1 Characteristics of two useful amino acid sets.

Signal class	Amino acids	#Signals of $x^2 > 10$	Pattern distribution
1	ARQELKM (SEQ ID NO:5)	84	clustering
2	CILMFWYV (SEQ ID NO:5)	216	anticlustering

Please add the paper copy of the sequence listing enclosed herewith, pages 1-2, at the end of the application.